[illegible]

NASA Grant NAG 2-123*

PILOT INTERACTION WITH AUTOMATED AIRBORNE DECISION MAKING SYSTEMS

FINAL REPORT

John M. Hammer, Principal Investigator

Center for Man-Machine Systems Research
Georgia Institute of Technology
Atlanta, Georgia 30332

*The NASA Technical Officer for this grant is Everett Palmer, NASA Ames Research Center, MS 239-2, Moffett Field, CA 94035.

Introduction

This research investigated ways in which computers can aid the decision making of an human operator of an aerospace system. The approach taken is to aid rather than replace the human operator, because operational experience has shown that humans can enhance the effectiveness of systems. As systems become more automated, the role of the operator has shifted to that of a manager and problem solver. This shift has created the research area of how to aid the human in this role.

The remainder of this report describes published research in four areas. It concludes with a discussion of the DC-8 flight simulator at Georgia Tech.

Archival Literature

1. Lewis, C.M., and Hammer, J.M. (1986). Significance testing of rules in rule-based models of human problem solving. IEEE Transactions on Systems, Man, and Cybernetics, SMC-16(1).

2. Yoon, W.C., and Hammer, J.M. (1988). Aiding the operator during novel fault diagnosis. IEEE Transactions on Systems, Man, and Cybernetics, SMC-18(1).

3. Yoon, W.C., and Hammer, J.M. (1987). A deep reasoning aid for deep-reasoning fault diagnosis. In G. Salvendy (Ed.), Human-Computer Interaction, Vol. 2, Amsterdam: Elsevier.

4. Yoon, W.C., and Hammer, J.M. (1988). Deep-reasoning fault diagnosis: An aid and a model. To appear in IEEE Transactions on Systems, Man, and Cybernetics.

Conference Literature and Technical Reports

5. Knaeuper, A., and Morris, N.M. (1984). A model-based approach for online aiding and training in process control. Proceedings of the 1984 International Conference on Systems, Man, and Cybernetics. Halifax, Nova Scotia, Canada.

6. Lewis, C.M. (1985). Rule-based analysis of pilot decisions. Proceedings of the Human Factors Society 29th Annual Meeting.

7. Lewis, C.M. (1985). Issues in rule identification and logical induction. Proceedings of the 1985 IEEE International Conference on Systems, Man, and Cybernetics.

8. Lewis, C.M. (1986). Identification of rule-based models. Unpublished Ph.D. dissertation. Atlanta, GA: Georgia Institute of Technology.

9. Rouse, W.B., and Morris, N.M. (1985). Understanding and avoiding potential problems in implementing automation. Proceedings of the 1985 International Conference on Systems, Man, and Cybernetics.

10. Yoon, W.C., and Hammer, J.M. (1985). Aiding the operator during novel fault diagnosis. Proceedings of the 1985 International Conference on Systems, Man, and Cybernetics.

11. Yoon, W.C. (1987). Aiding the Operator During Novel Fault Diagnosis. Unpublished Ph.D. dissertation. Altanta, GA: Georgia Institute of Technology.

Model-Based Online Aiding [5]

This research addressed the feasibility of adapting an existing rule-based system as an online "coach" for controlling PLANT, a simulation of a generic process plant. KARL, a rule-based model capable of controlling PLANT, was adapted to provide three types of information to subjects:

1) situation assessment (i.e., which operational procedure, if any, was applicable for a given situation);
2) guidance in following procedures (i.e., feedback whenever subjects' actions were inconsistent with available procedures); and
3) performance feedback (based upon changes in the system's stability).

Subjects received this information online while controlling PLANT. Compared to subjects in an earlier experiment who controlled PLANT without the benefit of the coach, these subjects maintained a generally more stable system, scored higher on a paper-and-pencil test of system knowledge, and were more successful in diagnosing an unfamiliar failure of the PLANT safety system. Careful analysis of these results in light of previous research with PLANT indicated that the reasons for these differences were not as straightforward as they might appear. This experiment is viewed as illustrating potential benefits and subtleties of using a rule-based model as an online coach.

Significance Testing of Rule-Based Models [1]

Many researchers have used rule-based systems to model human problem solving. Typically, the rule-based system has a large number of rules, each of which has several free variables that were adjusted during the modeling process. For the most part, significance testing of these rules has not been much of a consideration, although it should be. It is possible to describe N data perfectly with N rules using a trivial model that simply reproduces the data. While there is no evidence that this has happened in any of the research reported to date, there is a certain danger of overfitting a rule-based model.

Three methods were developed for testing the statistical significance of rules and other components of rule-based models. It was assumed that the percentage of behavior matched (e.g., commands) was the performance measure of interest. Two of the testing approaches, however, were not limited to this measure. They may be used to study any performance measure, though it may be possible for a rule to produce a statistically significant effect on one performance measure but not another. Rule testing by analysis of variance, randomization, and contingency tables was studied, and comparisons between these methods were developed.

Identification of Rule-Based Models of Problem Solving [6, 7]

Rule-based models have frequently been used to model human performance and behavior. A machine learning program was used to identify the rules employed by humans in two settings. The first setting was a collision avoidance maneuver for which the pilots had a cockpit display of traffic information (CDTI). This data was generated from an experiment to evaluate the effects of various CDTI displays on avoidance behavior.

The rules produced by the machine learning program can be combined in a decision sequence that accounts for a substantial portion of the maneuvers. When the intruder was maintaining a constant altitude, pilots executed vertical away maneuvers even for intruders posing no threat. This is the easiest of the maneuver decisions because it entails no geometric complications and was used whenever possible. For intruders changing altitude, a minority of pilots consistently checked for a threatening separation and remained on course if none existed. Another subgroup responded to horizontal threats by uniformly turning toward the intruder. This is a good decision if the intruder would have passed in front but aggravates the situation for intruders which would pass behind. The remainder of the pilots included this qualification in their decisions to turn toward the intruder. The mirror of this strategy, turning away from intruders which would pass behind was not observed.

The second setting was PLANT [Morris, N.M., and Rouse, W.B. (1985). "The effects of type of knowledge upon human problem solving in a process control task." IEEE Transactions on Systems, Man, and Cybernetics, SMC-15(6).], a simulated industrial process in which feedstock is pumped in at one end and the finished product is pumped out at the other. A three-by-three matrix of tanks connects PLANT input to output. Each tank is connected by valves to all tanks in adjacent columns. The operator controls valve positions and pumping rates for feedstock and product. Fluid dynamics are modeled within the system causing lags and oscillations to result when valves change state, as well as varying rates of flow due to relative tank heights. Valves trip closed when flow exceeds their setpoints. Failures of pumps and valves are also possible. The CRT system display shows tanks, their levels of fluid, open valves connecting the tanks, and numerical labels showing pumping rates and tank levels.

In concert these features produce a complex symbolic task in which conflicting goals relating production, system stability, long term trends, failures, and trips must be balanced to operate the system. At peak efficiency, all valves should be open, tank levels uniform across the system, and identically high pumping rates set for feedstock and product. PLANT is operated by subjects through a services of iterations which a control action is entered and the resultant updated system state displayed. The iterations from an experimental session (~500) provide a series of "snapshots" isolating specific system states and the responses subjects made to them.

In an initial analysis of this data [8], small sets of high coverage rules were assembled. Cross-validation was used to assess the reliability of the selected rules. Identified rules correctly matched 51% of control decisions in the identification sample for subjects in the control group and 32% of the control decisions in the validation sample. For subjects using PLANT procedures, combining symbolic (rule-based) and signal (internal dynamic model of PLANT) processing fared better matching control decisions 52% of the time. The generality of the well-performing rules obtained prohibited the detailed analysis of strategy possible in the CDTI case.

Deep Reasoning Fault Diagnosis [2, 3, 4, 9, 11]

This research studied the design and evaluation of knowledge-based aiding for a human operator who must diagnose a novel fault in a dynamic, physical system. Since the operator must employ deep reasoning about system behavior to diagnose such a fault, his or her performance may be restricted by cognitive limitations and biases. A computer aid based on a qualitative model of the system was built to help the operator overcome some of these limitations. This aid differs from most expert systems in that it operates at several levels of interaction which are believed to be more suitable for deep reasoning.

Four aiding approaches, each of which provided unique information to the operator, were evaluated. The aiding features were designed to help the human's causal reasoning about the system in predicting normal system behavior (N aiding), integrating observations into actual system behavior (O aiding), finding discrepancies between th two (O-N aiding), or finding discrepancies between observed behavior and hypothetical behavior (O-H aiding). Three experiments were conducted to evaluate the aiding approaches and to investigate the nature of deep-reasoning diagnosis. Human diagnostic performance improved by almost a factor of two with O aiding and O-N aiding. The results from the experiments were integrated into a model of human information processing in causal reasoning diagnosis.

DC-8 Flight Simulator

The failure to both complete and utilize the DC-8 flight simulator is a disappointment. An assessment of the cost of developing the simulation should have been prepared initially. The development breaks down into three categories: hardware, flight simulation, and display generation. The hardware category was completed at a cost of roughly $75,000. The flight simulation code is roughly one half done, and perhaps another 10,000 lines of code need to be written and tested. This would require one programmer-year to produce ($50,000). Display generation would require $15,000 in hardware and another programmer-year ($50,000). A total estimated cost of $190,000 compares favorably with the cost of a commercial product. However, the research funding needed to support such a facility must be larger than a single $100,000/year grant.

DEEP-REASONING FAULT DIAGNOSIS: AN AID AND A MODEL

Wan C. Yoon and John M. Hammer

Center for Man-Machine Systems Research

Georgia Institute of Technology

Atlanta, Georgia 30332

ABSTRACT

The design and evaluation are presented for knowledge-based aiding for a human operator who must diagnose a novel fault in a dynamic, physical system. Since the operator must employ deep reasoning about system behavior to diagnose such a fault, the performance may be restricted by cognitive limitations and biases. A computer aid based on a qualitative model of the system was built to help the operator overcome some of his/her cognitive limitations. This aid differs from most expert systems in that it operates at several levels of interaction which are believed to be more suitable for deep reasoning.

Four aiding approaches, each of which provided unique information to the operator, were evaluated. The aiding features were designed to help the human′s causal reasoning about the system in predicting normal system behavior (N aiding), integrating observations into actual system behavior (O aiding), finding discrepancies between the two (O-N aiding), or finding discrepancies between observed behavior and hypothetical behavior (O-H aiding). Three experiments were conducted to evaluate the aiding approaches and to investigate the nature of deep-reasoning diagnosis. Human diagnostic performance improved by almost a factor of two with O aiding and O-N aiding. The results

from the experiments were integrated into a model of human information processing in causal reasoning diagnosis.

INTRODUCTION

Becoming more of a monitor and supervisor in today's highly automated systems [Rasmussen 1984], the human operator must at times be involved in the task of diagnosing system failures, which is increasingly difficult as the system becomes more complicated and automated. The prevalent approach to fault diagnosis is to train the operator to have better knowledge and experience with commonly expected faults. The training might teach the operator to use symptoms to distinguish faults and to follow procedures to correct them. While this approach should be successful with common faults, it does not support diagnosis of novel faults.

Another, more recent approach is to support the human operator via expert systems for diagnosis. Those expert systems are typically based on a large collection of diagnostic rules, which associate symptoms to causes and generate tests. As for novel failures, many expert systems for diagnosis [Shortliffe 1976, Miller, Pople, and Myers 1984] are based on shallow reasoning: a set of symptoms suggests a diagnosis. This mapping is based on experience rather than a system model. Consequently, such systems are subject to the same limitations as training and procedures. The expert system designer has to anticipate the failure for the expert system to solve it correctly.

Aiding Based on a System Model

To diagnose an unanticipated, unexperienced fault, the operator must rely on his/her understanding of causality of the system [Davis 1984]. Such causal reasoning is usually a very demanding cognitive task when the system is complex. Therefore, an intelligent aid should be able to support the operator in causal reasoning about the system behavior. The most obvious way to achieve this is to let the aid run its own causal model of the system and provide the results to the human. A qualitative model of the system can be useful for this purpose.

Another advantage of an aid based on a causal model is that it should be more reliable and robust. The system knowledge is represented at the component level. Because components are small and comprehendable, it should be possible to create representations that are correct, perhaps even provably so. A system fault can be expressed as a combination of component faults which does not require a priori identification of the system fault itself. Thus, an aid based on a causal system model can cover a wider range of faults.

In spite of the power of the intelligent aid, we believe there are several reasons to keep the human in command of the problem solving. First, the current trend of automatic diagnosis is based on large rule-bases which are less useful in novel fault diagnosis. Second, the human and the aid may be better able to find a solution cooperatively than either can alone. This is possible, even necessary, because the human has better pattern recognition capabilities and can make inductive leaps. Third, in many cases, diagnosis is one of the subgoals and may interfere with other subgoals. For example, when diagnosis involves operating the system (e.g., opening valves, starting motors), it may interfere with the subgoal of system safety. The human is better suited for the responsibility of resolving tradeoffs in pursuit of an overall goal. Lastly, the human may need to resolve ambiguities inherent in the aid's model or even to extend the model.

Suboptimalities in Human Problem Solving

The aid is designed to mitigate human suboptimalities that occur during decision-making and troubleshooting [Wickens 1984]. Two categories of suboptimalities used here are knowledge-limited and cognition-limited. The knowledge-limited suboptimality is simply that the operator does not fully understand the system. Obviously, the aid's model is a basis for compensating for this problem.

Cognition-limited suboptimalities are of more interest when the system fault is novel rather than common. Novel fault diagnosis requires causal reasoning about the system, which is a cognitively very demanding task. The operator should repeatedly run a mental model of the system in multiple modes as well as maintain a diagnostic procedure. The required information processing can overload the operator's limited mental resources, especially attention and working memory. The results may be incorrect reasoning or inefficient use of information.

To help, the computer aid can process and display useful information so that the operator can use it. This may improve the system performance in two ways. First, the operator can dynamically allocate some subtasks to the aid and concentrate on others. This leads to lessened mental workload and improved performance on those subtasks undertaken by the operator. Second, since the aid reasons in parallel with the human, the human can confirm his/her results against the aid´s results. When the human overlooks some useful information or is affected by some biases, discrepancies would be noticed between the aid´s results and the operator´s own. The operator may then adopt the aid´s result to be used in subsequent reasoning. For example, when the human and the aid evaluate a hypothesis, the confirmation bias (i.e., the tendency to seek only confirming evidences) will be prevented since the aid, being not susceptible to this bias, would report disconfirming evidence.

Research Questions

It is likely that not every plausible form of aiding will improve operator performance. Some information which is both relevant and helpful may not be able to improve human performance because the human fails to incorporate the information into his/her problem solving. This leads to another question: which types of information are easily usable by the human? Our approach to answering these questions was, first, to build an aid based on the best principles available to us, and let the aid supply prospective types of information in experimental settings to evaluate their actual aiding effects. Successful and unsuccessful aiding may also provide insight on the architecture of human information processing.

In the subsequent sections of this article, we will discuss the suitable form of interaction for a deep-reasoning aid, the system which served as the context of problem, qualitative modeling of the system, the features of the aid, the experiments and results, and a model of human information processing in causal reasoning diagnosis. Because a literature review was included in recently published, early report of this research [Yoon and Hammer 1987], no review appears here.

LEVELS OF INTERACTION

In the design of interaction between the aid and the human, it is important to consider the nature of task to be aided. Deep-reasoning diagnosis has many subprocesses of which even the problem solver may not be aware. The aid should be able to help the human's processing without disturbing or interfering with it.

To discuss appropriate forms of interaction in this situation, we stratify the ways in which the human and computer interact into five levels in terms of intrusiveness (Figure 1). The two extreme (i.e., the most intrusive) levels are the human-direct level and the computer-direct level. In the middle, the human-suggest and the computer-suggest levels allow a problem solver, the human or the aid, to be moderately intrusive. Finally, there is the independent level at which neither problem solver influences the other. This stratification is orthogonal to the levels of required intelligence or knowledge the aid should have [Greenstein 1980].

At the _human-direct level_, the human assigns tasks to the computer. For example, the computer will respond to the operator's request to perform a subtask or to answer a question. The situation is opposite at the _computer-direct level_; the computer asks the human for some information or to perform some tasks. The human does not have a choice other than to follow the request.

Typical expert systems use only these two levels of interaction; some systems use only one of the two, others use both. At either level, the overall processing is serial and requires explicit communication. Certainly, this property does not promote the human's deep reasoning. The difficulty of human-direct level interaction is that the effectiveness of the aid depends upon the ability of the human to decompose the overall task into modular subtasks [Wickens 1984]. On the other hand, at the computer-direct level, the human does not have the freedom to pursue his/her own processing. This would reduce the benefit to the system of having the human whose flexibility and inductive and pattern recognition capabilities are superior to those of automation.

At the human-suggest level of interaction, the human may impose constraints on the computer´s processing. Examples are adjusting weights of different criteria, modifying the computer´s intermediate results, or restricting the computer´s attention to some area in the problem space. However, the computer will continue its tasks without explicit assignment by the human; only the data or criteria are modified. The computer-suggest level allows the computer to provide some information or warning to the human. The human is free to attend or not depending on his/her assessment of situation. The operator may postpone a response until finishing a current line of reasoning; or, the computer can be completely ignored. Thus, the communication is allowed to be less explicit and more abstract. What becomes a critical issue is that the suggestions by the computer need to be compatible with the human´s reasoning process.

At the independent level, both problem solvers pursue their own problem solving procedures without influencing each other. This level is almost non-existent in conventional expert systems which employ only the two extreme levels. When the interaction occurs at the suggest levels, however, the independent level fills the intermissions between suggestions. While there is no interaction, both problem solvers may be highly active in their problem solving. At times, the deep-reasoning process needs to be supported by interruption-free independence.

We believe that the three middle levels should facilitate more adequate aiding to deep-reasoning tasks. At those levels, the processing is more parallel and both problem solvers have more freedom. Two human problem solvers would interact mostly at those levels; they would suggest, take comments and hints, or be silent. Using the three levels of interaction was one of our principles in building the aid for novel fault diagnosis. Another related principle was to consider compatibility of aiding information with human information processing.

THE SYSTEM AND THE TASK

The Orbital Refueling System (ORS), a NASA-designed payload on the Space Shuttle, was selected for study [NASA 1985]. The function of the ORS is to refuel orbiting satellites with hydrazine, with the objective of extending their useful service life. As shown in Figure 2, the ORS fluid system contains a variety of components such as tanks, valves, pipes, etc. The operator controls the simulated ORS by opening and closing valves. Transferring fuel from propellant tank 1 to propellant tank 2 might proceed as follows. First, tank 2 pressure is reduced by momentarily opening valves 10, 11, 13, and 17. Second, tank 1 is pressurized by opening valves 1, 3, and 7. Gaseous nitrogen will flow out of the two small supply tanks, be pressure regulated, and fill tank 1 on one side of the bladder. To transfer fuel to tank 2, valves 5, 14, 15, 16, and 9 would be opened. Because this version of the ORS was for demonstration purposes, all transfers take place between the two large tanks rather than to a satellite fuel tank. There are several assemblies whose purpose was not explained in the above example. The relief valves RV1 and RV2 serve as a safety pressure relief. Check valve CV1 prevents backflow into the gas system. The bladders in tank 1 and 2 serve to isolate the fuel from the propellant and also to contain the fuel in the weightlessness of space. Some components (e.g., valves 10 and 11) may seem redundant; they are so by design for two failure tolerance.

Nomenclature

In discussing the ORS and the operator's actions and diagnosis, we have found the following nomenclature useful. A component is the smallest unit of the ORS system that is modeled in isolation. Typical components include valves, tanks, pipes, regulators, sensors, etc. The entire set of components, working together according to the qualitative dynamics, is a system. A path is a connected set of components, which could be either a graph-theoretic path or tree.

Components have states. For example, a valve may be open, closed, or leaking. The state is what the component is actually doing. A commanded state is the state to which a commandable component asked to assume. For example, a valve may be commanded open or closed. A component also has a behavior mode, such as fail-open or normal. The behavior mode describes the states which the component takes in response to commands and external conditions. For example, a fail-open valve is always open, regardless of the command.

The Diagnosis Task

The operator's task is to diagnose the failure in the system. This requires the operator to manipulate and observe the system, because a diagnosis cannot be determined uniquely from an observation of a state vector at a single point in time. A solution is an assignment of states to components such that the assignment's behavior is always identical to system behavior. For a single valve failure, the solution would be a normal state for all components save the failed valve, which might be jammed shut. The diagnosis problem can be viewed as a combinatorial search for a state assignment. The search is constrained by the laws of component physics. That is, a state assignment to a component imposes constraints on its neighboring components. For example, if a valve is opened and permits a flow down a pipe, the component receiving the flow must be in a state to accept the flow.

QUALITATIVE MODELS OF CONTINUOUS PHYSICAL PROCESSES

This section describes qualitative models: representations, the computational problems solved, and the specific needs of our aid of the qualitative model.

A qualitative model is a symbolic representation of a system. Its most basic description is of a component. A component is described in terms of its connections to other components and its behavior. Behavior is described in terms of the physical variables which are present at its connections. The differentiation between the structural description (connections) and the behavioral description is particularly important for insuring the robustness of a qualitative model. The isolation of each component in the behavioral description has usually been emphasized by other qualitative modeling [De Kleer and Brown 1983]. Contrarily, our qualitative model represents the system at both the component level and at an aggregated level as paths. The motivation for this is the belief that a multi-level description is closer to the operator's internal model of the process. In fact, more effective communication between our model and the human operator was enabled by the use of the higher level description.

From a given state, the behavior of a component is described in terms of the physical variables present at its ports. A physical variable (and its time derivative) may take several values. The time derivative usually has only one of three possible values: negative, zero, or positive. The variable itself may take either nominal or ordinal values. The nominal values usually correspond to points at which behavior (component or material) changes. For example, water temperature would have nominal values at freezing and boiling. Variables may also take on ordinal values (or relationships). For example, water temperature could be taken to be greater than freezing and less than boiling.

The nominal and ordinal values taken by physical variables are said to occur in a quantity space [Forbus 1984, Kuipers 1984]. The quantity space is a partial ordering on the physical variable values it contains. The partial ordering occurs because not all comparisons are relevant to understanding the physical system qualitatively. For example, consider a valve between two tanks, A and B. When the valve is opened, the resulting behavior is determined by the pressures in two tanks. The pressure at other unconnected points in the system is unrelated to the above behavior.

AIDING WITH A QUALITATIVE MODEL

This section describes how a qualitative model is used as a foundation for aiding. First, each window of the interface will be described. Four different aiding strategies and the motivation for each of them will then be presented. Each strategy emphasizes different type of aiding information.

ORS Interface

The interface has four windows: schematic, interaction, sensor display, and hypotheses (Figure 3). The schematic window displays a schematic diagram of the ORS. The schematic always shows the commanded state of the valves. The interaction window is where the operator's commands are echoed by the interface. The commands available to the operator include the following:

(1) Opening and closing valves.

(2) Comparing two pressures. On a real physical system, the numerical pressure could be displayed on the schematic. When a qualitative model is used to simulate the physical system, there is no absolute scale in general to which a pressure can be referred. Instead, a pressure can be compared to other pressures in the system by the relations less-than, equal-to, or greater-than.

(3) Display of the first derivative of a pressure (positive, zero, or negative).

And, when the corresponding aiding feature (it is described more fully in a later section) is available,

(4) Turning the what-if model on and off.

(5) Making state assumptions in the what-if model.

The sensor display contains the output from the sensor display commands: the relationship between two pressures or the first derivative of a pressure. When appropriate aiding features are activated, suggested sensor readings will also be displayed in this window.

The hypotheses window displays a set of hypotheses that are set by the operator. These hypotheses are simply state assignments to components (e.g., valve 13: leaking). Pipes, which do not have names displayed in the schematic, are designated as left or right to named components such as valves and orifices. For example, the pipe between valve 8 and orifice 4 is designated either R V8 or L O4.

Aiding Approaches

Based on observed human strategies of diagnosis, four aiding approaches seemed to deserve evaluation. Each approach emphasizes different information and uses an appropriate communication mode for the kind of information.

Topographic Aiding. The first and second aiding approaches are based on two presumed forms of operator cognitive processing. First, the operator must observe and infer what the system is actually doing. This processing is termed O (Observed) and is concerned with flows, leaks through valves, leaks out of pipes, and the general vicinity of the fault. Second, the operator needs to generate normal system behavior to compare with observed behavior. This processing is termed N (Normal). Two obvious forms of aiding are to generate O and N so that the operator does not have to devote cognitive processing to generating them. To produce O, the aid integrates the information from the pressure sensors to which it has continuous access. Like a human operator, the aid has to guess the actual behavior from the sensor information since it does not know the real system state. In contrast, N is generated by the qualitative model under the assumption that every component is in the normal behavior mode.

O and N are displayed topographically. For both O and N, the aid displays two forms of system behavior: equal pressure paths and mass flow paths. The former is the set of components that should be at equal pressure given the commanded valve positions. Whenever the operator creates an equal pressure path by opening a valve, the path is highlighted. Similarly, a mass flow path created by an operation is highlighted as long as it exists.

Figure 4 is an example of N display. Opening valve 9 was the latest change. This would make, if the system were fault-free, the pressure is equal through the highlighted path.

Figure 5 shows the same configuration as Figure 4, except that the O display (rather than N) is activated. Suppose that when valve 9 was opened, the pressure P2 began to decrease and P1 increase. This leads the aid to believe there is a mass flow from tank 2 to tank 1 (the path is highlighted) in spite of the closed positions of valve 8 and valve 15. However, since the aid cannot be certain which valve is leaking, it highlights both paths. When a precise conjecture is not possible, the aid will take a conservative position as in this example. Note that O and N aiding cannot be used simultaneously.

<u>Differencing Observed and Normal Behavior</u>. The third aiding approach is to suggest observations that reveal the differences between the observed system behavior and the normal system behavior. This difference will be referred to as O-N. The importance of O-N in ORS diagnosis was discussed in connection with the results of our preliminary experiment [Yoon and Hammer 1987]. Such a deviation from normal behavior, when observed and correctly interpreted, helped effectively reduce the size of the feasible hypothesis

set. Figure 6 shows an example of this feature's display in the same situation as of Figure 4 and 5. The aid suggests, for example, to issue a command *D P1*, which is to inquire the first derivative of P1. When the operator follows this, he/she will find P1 is increasing, which is opposite to the commanded situation (no flow should be possible from either GTK or TK2G/L).

The What-if Model. The fourth, and the last, aiding feature is closely related to the above. This feature can use any hypothetical behavior (denoted by H), instead of the normal behavior, with which to difference the observed system behavior. The operator can freely set or remove hypotheses. Then, the aid will run a what-if model based on the hypotheses in place of the normal model. Any discrepancies (denoted by O-H) will be reported in the same way. If the hypothesis is incorrect and the observed and hypothesized bevavior differ, the aid will recommend readings that indicate the difference. If the hypothesis is correct, the aid will produce no recommendations. For example, suppose valve 8 is leaking to allow a flow from tank 2 to tank 1. If the operator's hypothesis is a leak in the pipe between valve 10 and 11, the feature would present a display shown in Figure 7. If the hypothesis were right, P1 should not increase. In this example, P1 does increase, so the aid recommends a reading *D P1*. Also, the hypothesis does not explain the difference between P2 and P4. Note that if no hypothesis is stated, the recommendations would be the same as the previous example (i.e., O-H = O-N if H = N).

The common motivation for these aiding approaches is to perform computations that the operator is believed to do when diagnosing the system. As much as these computations are related to the human's mental model, the qualitative model in the aid may be an appropriate vehicle to help or replace the computations. There are two ways this approach might help. First, the operator may have an incorrect or incomplete mental model. Second, the operator may have difficulty integrating correct component behavior into correct system behavior because of cognitive limitations. The aiding approaches support different uses of the mental model: to envision the normal or hypothetical behavior, to conjecture the actual behavior, and to describe the difference between behaviors of two (e.g., O and H) models. This does not mean the operator need not understand the system at all; he or she still needs to understand the meaning of aid's information and select the hypotheses.

THE EXPERIMENTS

Overview of Experimental Design

To evaluate the types of aiding information, three separate experiments were conducted. The first experiment tested the effects of N information. The next experiment compared the effects of O and O-N against unaided diagnosis. The last experiment focused on hypothesis testing and evaluated the aiding effects of O-N and O-H.

The display of aiding information prevented those features from being tested together. A subject must not be exposed to both N and O features since severe interference, perhaps in the form of a carry-over effect, was expected. This is because the display of O and N information is identical but each carries a different meaning. O-H and O-N for the same reason should not be used together. When O-H is used, it acts as O-N until the subject expresses one or more hypotheses. This makes a direct comparison between O-N and O-H difficult. Even if O-H really improves the performance, its contribution will be depend on the extent to which a subject uses it. Therefore, a different experimental setting needs to be employed to evaluate the potential benefit of O-H. The above considerations led to the three separate experiments.

In all three experiments, replicated Latin square designs were employed [Edwards 1972]. Differences in the complexity of problems and differences between users were expected to introduce large variation in the performance. It was therefore desirable, in order to enhance the efficiency of the experiments, to select problem and subject as two blocking variables. Such designs are called within-subjects designs for each subject serves in more than one treatment level.

A Latin square design, if its assumptions hold, should be more economical than a corresponding complete block design. Even without considering economy, our experiment does not allow a complete block design. Because a subject should not be given a same problem more than once, he/she can be assigned only one level of treatment for each problem.

In a Latin square design, the positions of each treatment level are counterbalanced: namely, each treatment occurs at each test position with equal frequency. This prevents possible practice effects from being confounded with treatment effects. Instead, practice effects are then confounded with test positions (i.e., problem). However, the problem factor is merely a blocking variable and we were not interested in the significance of its effects. Also, the training was designed to stabilize the subject's performance and thus minimize learning effects.

One possible problem with a within-subject design is that the value of an observation for one treatment may be influenced by the effects of treatments applied during earlier periods. When this arises, the treatment is referred to as having carry-over effects. The influence of this effect, if any, may be partially compensated for by adopting a balanced Latin square design, in which each treatment follows every other treatment the same number of times. When the number of treatments is odd, then at least two Latin squares are required to achieve this. This replication also permits a larger number of data points. All our experiments were designed following the above principles. The resulting designs are presented in Figures 8, 9 and 10.

While the balanced Latin square designs may compensate for the above problems, they are based on several assumptions. A key question concerning the Latin square design model is whether the effects of blocking variables and treatments are additive: since there is only one observation per cell, a Latin square design model assumes additivity to estimate the error variance. If nonadditivity is present in the data, the use of a model assuming additivity will lower both the significance level and the power of the test for treatment effects. Thus, the Tukey test for additivity was conducted whenever we applied a model to the data [Neter and Wasserman 1974, pp.780].

While homogeneity and normality of error variance are the basic assumptions in an ANOVA model, it is known that the F test is not much affected by deviation from these conditions [Lee 1975, pp.284]. However, a residual plot of error terms against expected cell means can reveal the need for transformation of dependent variables. Since a transformation would affect the interpretation of treatment effects, residual plots were examined in every analysis.

Experiment 1

The purpose of this experiment was to compare N aided and unaided diagnosis. It is reasonable to expect diagnostic performance to be improved when the envisionment of normal system behavior is improved. In our pre-experimental observations, however, we observed that most subjects found this aiding confusing or irrelevant. Since its effectiveness was doubtful based on this observation, it was evaluated first.

Six industrial engineering students volunteered to serve as subjects. They were trained through two sessions (total 3.5 ~ 4.0 hours) to acquire enough knowledge of fluid dynamics and elements of diagnostic procedure. The goal of our training was to teach the subjects correct causal reasoning about the ORS and give them reasonably stabilized diagnostic skills. However, if a subject is exposed to a kind of problem several times in a short period, the subject may develop some mechanistic diagnosis procedures that do not require causal reasoning. When a similar problem is given, the subjects may try to deal with it as a routine failure rather than a novel one. We felt that a longer training may increase this possibility since the complexity of our version of ORS is only moderate.

Training session 1 started with basic principles derived from fluid dynamics. Then, possible malfunctions for each component were discussed. Finally, the subjects undertook a simulated ORS mission, during which envisioning of normal system response was practiced. Session 2 taught elementary diagnostic procedures such as checking a sensor bias or a valve leak. The subject then was required to plan testing procedures for five typical hypotheses. Each developed procedure was discussed by the experimenter until the subject developed (and understood) a correct procedure. Next, three real problems were given as exercises. Sessions 1 and 2 took 1.5 hours and 2 hours on the average, respectively.

The performance of the subject in the entire training sessions was closely monitored. The first session contained many questions to ascertain if the subject achieved proper understanding. The answers were checked during the same session and, whenever necessary, discussed again. Problem solving exercises were also attended by the experimenter and necessary discussion or re-explanation was provided. The result was that the initially poorer subjects would spend more time in training rather than end with poor understanding. By the end of the second session, all the subjects performed satisfactorily and showed little additional improvement in diagnostic skill.

The considerations which led to the design of experiment has been discussed in the overview section. The design for experiment 1 is shown in Figure 8. Each group was composed of three subjects and the Latin square was replicated three times using different problems.

Many different performance measures were tried with our data from the pilot experiment. The number of information gathering actions (#IGA) and the time to solve (Time) appeared to be appropriate performance measures. Although several other measures were examined with the data, they either turned out to have insufficient resolution or showed high correlations with the above measures. Thus, the above were the most important measures in this experiment. Time and #IGA showed virtually identical behavior both in the examination of aptness of the ANOVA model and tests of significance.

The data collected from 36 subject-problems were first analyzed to determine if there were significant interactions between problems and aiding levels. The interactions were found insignificant both in time ($p = .409$)

and #IGA (p = .534). This suggested that the interaction term can be excluded from the model and its sum of squares may be pooled with that of error term.

The Tukey test uncovered nonadditivity in the data of both Time and #IGA. The residual plot indicated that the cell standard deviations were proportional to cell averages. As this is frequently the case when the criterion is response time [Lee 1975, pp.291], a logarithmic transformation was suggested. After the transformation, the anomaly in the residual plot was fixed. The transformed data, both in Time and #IGA, appeared to adhere to the homogeneity and normality requirements for ANOVA better than the original scores. The interactions between aiding levels and problems were still insignificant. The Tukey test was performed again with the new scores and showed no significant nonadditivity.

The contribution of N aiding to both Time and #IGA was on the negative side, though not significant (p = .096 and .381, respectively). On the average, it corresponds to 31% increase in Time and 13% in #IGA.

These results may not simply be interpreted that N feature did not help the envisionment of normal system behavior or that the role of such envisionment in the diagnosis is unimportant. A proper interpretation may be that the normal envisionment could not be helped very well by providing external information because the process is too quick and deeply embedded in a larger cycle of human information processing. Another possibility is that envisioning normal system behavior was not a bottleneck in diagnostic performance.

We concluded the former interpretation was very likely considering the following. First, most subjects, after their main sessions, stated that the aid was not only uninformative, but also somewhat distracting or confusing. A subject said he wished he could get ´real´ system behavior rather than ´normal´ behavior. Second, the fairly strong negative aiding effects could not be explained if the aid helped only unimportant subtasks. Third, the negative aiding effect was notably stronger in Time than in #IGA. (This was the only occasion in which the two measures showed any notable difference in the analysis throughout experiment 1 and 2.) This implies that the aid forced the subjects to think for a longer time but did not greatly affect their diagnostic procedure. This result supported the subjects in reporting that the aid was confusing and distracting. Thus, we concluded that there was interference between N information and the operators´ diagnostic information processing. Certainly, they do predict normal system behavior as a subtask: it is obviously necessary. But, when they seek information from the display, it was not of normal system behavior. This observation will be implemented in modeling of deep-reasoning diagnosis later in this paper.

Experiment 2

The second experiment was to assess the aiding effects of O and O-N features against unaided diagnosis. Nine new subjects, again industrial engineering students, were recruited as volunteers. Two training sessions which were virtually same as in the first experiment were given. In terms of content, the only difference was that the explanation of the new features replaced that of N feature. The design of experiment, shown in Figure 9, was also the same except for a different number of treatment levels and replication.

The procedure of statistical analysis was the same as in Experiment 1. First, the interactions between aiding levels and problems were found insignificant. After pooling the sum of squares for interactions into error sum of squares, the Tukey test for additivity was performed. No significant nonadditivity either in Time or #IGA was found. When the residual plots were examined, however, it was indicated that both measures needed to be logarithmically transformed. After the transformation, the new residual plots showed stabilized error variance. Again, the interactions between aiding levels and problems were insignificant. The Tukey test with the new scores yielded a much lower F value than before the transformation, confirming that the new scores fit the assumptions of the model better.

As results of the analysis of variance, both Time (p = .0302) and #IGA (p = .0005) showed significant effects of aiding. In Time, the improvement (i.e., decrease in Time) on the average was 34% by O aiding and 42% by O-N aiding. In #IGA, O aiding permitted 40% decrease while O-N aiding gave 44%. Neuman-Keuls tests were performed to determine if there were significant differences between pairs of aiding levels. Both O and O-N aiding levels had significantly different means when compared to the unaided mean. This result was identical for both Time and #IGA. In any measure, there was no conspicuous difference between O and O-N aiding.

The obvious conclusion is that both aiding features were effective in both measures and permitted solid enhancement of human diagnostic performance. In contrast to the N feature, these types of information appeared to be well accepted by the human process of diagnosis and helped the human in some important elements for his/her performance.

Experiment 3

The motivation for Experiment 3 was informal observation of subjects during Experiment 2. The effectiveness of O-N aiding in Experiment 2 appeared to decrease as the diagnosis proceeded. As is to be supported by more elaborate analysis later, this motivated us to investigate possible transitions between problem solving phases made by the diagnostician. Probably the most notable change in diagnosis as time passes was that the diagnostician began to deal with more explicit and individual hypotheses after the feasible hypothesis set size had been sufficiently reduced. In later phases with individual hypotheses, the characteristics of problem solving may be very different than the earlier phase of narrowing down the hypothesis set. Therefore, it was necessary to investigate the nature of diagnostic activity and proper form of aiding with such explicit hypotheses.

Due to its unique purpose, this experiment had an important difference in its setting from the first two experiments. In Experiments 1 and 2, the subjects solved whole diagnosis problems starting with primary symptoms. In the third experiment, the subjects determined whether a given hypothesis was true. At first, instead of being told of symptoms, the subject was allowed to perform some predetermined sensor readings which would indicate abnormal system behavior. Then, the subject was given a hypothesis to evaluate. Without needing to diagnose the real failure, the subject was to end the problem solving merely saying if he/she agreed at the hypothesis.

The effects of O-N aiding and O-H aiding were evaluated against unaided situations in two separate Latin square designs, i.e., Experiments 3-a and 3-b. They are shown in Figure 10. This was because, as mentioned earlier, it was not possible to assign both O-N and O-H aiding levels to the same subject due to expected interference. Although both Time and #IGA were collected, only Time was used in formal statistical analysis. Since the problems are much smaller in size than those of earlier experiments, #IGA is usually a small integer that would not easily lend itself to meaningful statistical analysis considering the vast difference in the subjects' diagnostic procedures. Otherwise, the analysis proceeded in a similar procedure as that of previous experiments.

In the analysis of the data from Experiment 3-a, the main question was what effects O-N aiding will have on the performance of diagnosis with a given hypothesis. First, the interactions between aiding levels and problems were tested and found insignificant (p = .881). Thus, a pooled error sum of squares were used for subsequent analysis. The Tukey test for additivity revealed the data were indeed additive. The residual plot also confirmed the model fitted the data quite well. It may be noted that, unlike the former experiments, no transformation was found necessary. The reason perhaps lies in the nature of the problems; these problems are just elementary subtasks which the operator should do numerous times in a whole diagnosis. As for the whole diagnosis time, the standard deviations were proportional to the means. That is, when a problem was more complex, the variation in the actual diagnosis time tended to be larger. This tendency most probably comes from the process of narrowing down the hypothesis set since the subtask of hypothesis testing did not show this property.

The performance was somewhat worse with O-N aiding than without it. Although not significant (p = .192), the difference on the average extended to 15.6 seconds (overall average was 67.4 seconds). The interpretation will be discussed with the evaluation of O-H aiding.

Experiment 3-b proceeded the same way except O-H aiding was tried in the place of O-N aiding. The interactions between aiding levels and problems were negligible (p = .8593). The additivity test confirmed that the data were additive. As in Experiment 3-a, the residual plot indicated that no transformation was needed. surprisingly, the effects of aiding appeared to be completely negligible (around 1 second, p = .9546).

The interpretation of these results is subtle. First, the O-N information was not relevant to the operator's activity to test a given hypothesis. The aid distracted the operator only to think about irrelevant information. This confirmed our earlier observation in Experiment 2 that the aiding effects of O-N information seemed to diminish as the diagnosis proceeds into its final stage. This observation, too, became a basis of our modeling of deep-reasoning diagnosis which is discussed in a later section.

Then, why was O-H aiding, which must be relevant to the given hypothesis, not effective? Two possibilities occur. First, the O-H information was simply not relevant to the problem solving. Otherwise, the information was relevant but trivial to the subjects. The first interpretation is not consistent with our previous results that, when irrelevant information was given to the subjects, the performance showed signs of degradation. The remaining choice is that the information, which is basically a set of suggestions for interesting observation, was already known to the subjects. That is, they already knew what to see even without the aiding; the aid only confirms it.

This interpretation could be further confirmed by a detailed process analysis. In Experiments 3-a and 3-b, 32 problems were solved without aiding. If O-H aiding had been provided with these problems, it would have suggested useful sensor reading actions 39 times. In 38 out of the 39 times (97.4%), the subjects collected equivalent information without it being suggested. Since they were ready to gather the O-H information whenever it was useful, the suggestions for this information by the computer were not able to improve the performance further. Because, unlike the O-N suggestions, O-H suggestions were just what the subjects were about to do, they were understood as trivial so that no performance decrement was caused by interference, either.

There was also an indication that the subjects planned valve operations and sensor readings together ahead of the actual operations. The subjests' collecting of O-H information was remarkably precise. There were 5 occasions in which the O-H aiding, if had been given, would have suggested uninformative readings. Failing in only one case out of 39 to look at useful O-H information, the subjects did not waste their time to do the uninformative sensor readings in any of the 5 occasions. Such precision may not be possible if the subjects were simply hunting around for useful observations by chance in scenes they just created. Most likely, the scenes were purposely planned aiming at the useful information. It should be noted that this tendency was unique and appeared only when an explicit hypothesis was given.

Summary

To summarize, O aiding and O-N aiding improved the diagnosis while N aiding did not. Actually, N aiding seemed to have negative effects. This suggests that the operator can effectively utilize O information, not N information, supplied from outside of his/her own information processing.

The usefulness of O-N aiding seemed to decrease over time perhaps as explicit hypotheses arose. In explicit hypothesis testing, O-N aiding showed a weak negative contribution while O-H aiding did not affect the performance at all. When weak negative effects were found, there seemed to be some interference caused by irrelevant information. On the other hand, O-H aiding was trivial and innocuous. The precision with which the subjects collected O-H information indicated that, when a hypothesis was given, the operational actions and data collection were usually planned together before the operations. This is an important observation in how the operators used their mental models.

A MODEL OF DEEP-REASONING DIAGNOSIS

Methodology

In this section, the experimental results will be integrated into a model of novel fault diagnosis.

The overall diagnostic procedure can be viewed as a combination of two elements: information processing tasks and a control strategy. Information processing tasks are subprocedures of diagnosis which can be characterized by their input, output and processes which take the input to produce the output. The control strategy is the way in which information processing tasks are selected.

The emphasis in this research has been on the information processing tasks, not the control strategy. There are several reasons for this. First, aiding novel fault diagnosis is the goal. Such diagnosis relies on causal reasoning about the system. To help causal reasoning, information processing tasks in which causal reasoning is embedded need to be understood. Second, we wanted to evaluate an aid which would be able to help the human to overcome cognitive limitations by some extent. While the aid would possess a similar causal reasoning capability to a human, it would not suffer the same cognitive limitations. This aid would be a more direct help to information processing tasks rather than the control strategy. Third, the findings from our research would permit insights to the structure of these information processes since our aiding approach was to provide various types of information which would substitute for the operator's information processing.

The emphasis on information processing led to a description of data flows rather than a flow chart. A flow chart would depict how the chronological sequence of various processes is controlled. In contrast, a data flow diagram would describe the necessary information input to a process, the expected output from a process, and the organization of processes through the links of information. This diagram helps to identify necessary subprocesses and alternative ways of automation.

A basic assumption connects our aiding experiments and the human information processing model: the human can better incorporate external information into his/her processing when the information becomes an alternative input to one of the higher level processes. An information processing task can be broken into _processes_, each of which can be broken into _subprocesses_. We assume that aiding information can be substituted for an entire process more effectively than for just an individual subprocess. There are several reasons to believe this assumption is reasonable. Because they are inner cycles in processes, subprocesses iterate and require input at higher rates. Also, the operator's working memory is more heavily loaded during a subprocess since the status of the higher level process, as well as that of the subprocess itself, should be retained. With the frequent cycles and heavy mental workload, it would be harder to perceive and apprehend externally supplied information [Wickens 1984, Rasmussen 1984].

As far as causal reasoning of the system operation is concerned, two directions of information processing should exist: observations to hypotheses and hypotheses to observations. The former task takes observations as input and produces hypotheses, while the latter starts from hypotheses and identifies necessary observations. Both tasks may be categorized as search by evaluation according to Rasmussen's classification [Rasmussen 1984].

Observations to Hypotheses

This task is triggered by observations of system behavior and will be referred to as data-driven search. It occurs when the observations were collected without particular hypotheses or showed unexpected patterns that fell outside hypotheses of interest. It seemed therefore natural that the subjects performed this type of process more often in earlier phases of diagnosis. Since O-N aiding was useful in earlier phases, the information it supplied must be closely related to this task. The poor performance of N aiding, however, indicates that the human's use of N information is in a lower level subprocess, very likely to produce O-N information. Therefore, it is suggested that there is a process which filters the observations to pass only more interesting (i.e., unexpected) ones to the next process: N information is used for one of its subprocess. Obviously, there must be one more process to complete this task. In this second process, the human tries to come up with a set of plausible hypotheses that explain the observations.

Some of the interesting observations may be remembered to evaluate future hypotheses throughout the diagnosis. The above constraints allow one to conceive a model of the data-driven search as represented in Figure 11.

Two processes were identified. The first process is filtering observations. Only the observations which passed this filtering are used in the following process of entertaining hypotheses. The filtering process contains a reference mental model of the system. The reference model is a mental model that produces standard behavior against which observed system behavior is continuously compared and judged as expected or unexpected. At first, the reference model behavior is that of normal system. As more observations are accumulated, however, some abnormal system behavior would also become expected even though the reason may not be understood. An expected observation does not carry additional information and should be filtered out as trivial. Thus, the reference model should evolve incorporating more and more observations of actual system behavior. Converging to the actual system in its behavior, the reference model would lower the probability of unexpected observations. Consequently, the efficiency of unplanned observations would decrease and the data-driven search would become less useful as the diagnosis proceeds.

In earlier phases of diagnosis, when the reference model behavior is normal, O-N aiding replaces the whole filtering process and provides input information to the hypotheses entertaining process. According to our basic principle, it should be easier for the human to incorporate such information into his information processing. This was supported by the experimental result that O-N aiding improved the diagnostic performance. However, the gradual departure of the reference model from normal system behavior would degrade the relevance of O-N aiding in the filtering. It was supported by the observation that O-N aiding was mostly useful in earlier phases of problem solving.

O aiding enhanced the observations which are input to the filtering process. The enhancement is in fact presentation of observed system behavior at a higher level of abstraction than the sensor displays [Rasmussen 1984]. For example, while the operator would normally look at individual pressure points to check the system behavior, O aiding would display a mass flow which is not the behavior of a component, but of a path. Since this level, being more functional, allowed more appropriate information coding for the operator´s use, it should improve the filtering process. The experimental results supported this.

The prediction of normal system behavior (N aiding) is at first equivalent to the subprocess of running the reference model. This activity is internal to the filtering process, neither replacing a process nor providing better information to a process. As a result, there may be little chance to improve human diagnosis by providing this information from outside. Actually, the experiment showed that N aiding had rather negative effects, though not significant, perhaps due to distraction.

Hypotheses to Observations

When given hypotheses are to be evaluated, the operator would build a testing plan that may prove one hypothesis and disprove the rest. This task is called hypothesis-driven search. Experiment 3-a indicated that, by demonstrating poor performance of O-N aiding, this task was very different from data-driven search in its information processing.

This type of process tends to be employed more often toward the final stage of diagnosis as the data-driven search loses its efficiency. An important restriction of this process is that the hypothesis should be sufficiently explicit for the diagnostician to perform mental simulation based on it. There are usually too many explicit hypotheses that are feasible in earlier phases of diagnosis. Therefore, the data-driven search may be preferred in narrowing down the feasible hypothesis set. Toward the end of diagnosis, however, the number of feasible hypotheses would become smaller and the need of testing the remaining hypotheses individually would increase. Then, the hypothesis-driven search dominates the diagnosis.

In Experiment 3-b, we forced the subject to perform this process by assigning a hypothesis to test. The experimental result that O-H aiding did not improve the human diagnosis can be explained in this model. O-H aiding suggested sensor readings which would show the difference between actual and hypothetical system behavior. When the hypothesis is false, a right test would reveal the existence of O-H behavior to disprove the hypothesis. Thus, O-H information is certainly relevant to the hypothesis testing. It is reasonable to expect O-H aiding to be helpful if the operator collects observations and filters them as in the data-driven search. If, however, the tests are planned by predicting observable differences (as in Figure 12) depending on whether the hypothesis is true, O-H information is identified before the actual testing operation. In this case, externally suggested O-H information would only be redundant and would not improve the performance.

The latter case was supported by the experiment; the aid gave no performance improvement; the operators collected O-H information in an extremely efficient manner even in unaided diagnoses, in which they were not given the suggestions by the aid. Therefore, it is safe to conclude that the operator, when a hypothesis is given, runs his/her mental model to determine a test that would distinguish the given hypothesis from other hypotheses. Figure 12 describes the model of this task.

Control Strategy

The control strategy is both highly dynamic and individualistic. Operators switch frequently between information processing tasks. The selection of tasks depends on the assessment of relative efficiency and effectiveness of different tasks in different situations. For example, if the diagnostician is equipped with very inexpensive testing methods to check every component directly, the cost of hypothesis-driven search will be drastically reduced from what it is in the ORS diagnosis. This observation suggests the possibilty that the control strategy can be changed when aiding affects the efficiency of elementary tasks.

Although the two information processing tasks are the most important elements, the strategy may involve other types of information processing. Topographic search [Rasmussen 1984] can be used either to entertain hypotheses or the necessary observations for a hypothesis. In fact, this is believed to be the frequent way in which the operator, when performing data-driven search, selected the data to begin with.

Regarding the control strategy, the only observation we could be assured of was that the subjects gradually transitioned from data-driven to hypothesis-driven search as the diagnosis proceeded. This was perhaps because the reduction of the size of feasible hypothesis set changed the relative efficiency of two processes. For instance, with only one hypothesis to deal with, explicit planning of test by hypothesis-driven search must be more efficient. It may also be partly because, as we have already discussed, the data-driven search lost its efficiency as observations were accumulated.

As a conclusion, the detailed modeling of information processing tasks helped to integrate our findings and observations of human operator's novel fault diagnosis. The models of human information processing tasks were useful in explaining the aiding effects of various types of information. It should also be useful to predict effects of aiding to be proposed in the future. Such predictions, in turn, may be tested in experiments to verify the model.

CONCLUSION

An aiding approach has been described and evaluated for novel fault diagnosis in complex systems. To the best of our knowledge, this approach is unique in the following ways. First, the emphasis is on novel rather than routine faults. Second, it contains a qualitative model that may correspond to the human's internal model of the system. This model represents knowledge only of how the system behaves. Therefore, this aiding approach does not rely on proceduralized knowledge. Third, the qualitative model is the basis for much of the aiding that takes place.

The experimental results confirmed that a deep-reasoning diagnosis can be aided, without disturbing the human diagnostic procedure, by providing relevant information. However, the results also suggested that the aiding information should be compatible with the human information processing. This emphasizes the importance of understanding the human information processing to build an effective aid. A principle of particular importance is that the information from/to higher-level processes is better incorporated into the human's information processing. The findings and observations were integrated into an effort to model the information processing tasks for deep-reasoning diagnosis.

ACKNOWLEDGMENT

This research was supported by NASA-Ames Research Grant NAG 2-123. Dr. Everett Palmer is the technical monitor.

REFERENCES

Davis, R., "Diagnostic reasoning based on structure and behavior," Artificial Intelligence, Vol. 24, pp. 347-410, 1984.

De Kleer, J. and Brown, J.S., "Assumptions and ambiguities in mechanistic mental models," in D. Gentner and A.L. Stevens (Eds.), Mental Models, Hillsdale, NJ: Lawrence Erlbaum Assoc., 1983.

Edwards, A.L., Experimental Design in Psychological Research, New York, NY: Holt, Reinhart and Winston, 1972.

Forbus, K., "Qualitative process theory," Artificial Intelligence, Vol. 24, pp. 85-168, 1984.

Gentner, D. and Stevens, A.L. (Eds.), Mental Models, Hillsdale, NJ: Lawrence Erlbaum Assoc., 1983.

Greenstein, J.S., "The use of models of human decision making to enhance human-computer interaction", Proc. of 1980 IEEE Int'l Conference on

Cybernetics and Society, Cambridge, October, pp. 968-970.

Kuipers, B., "Commonsense reasoning about causality: Deriving behavior from structure," Artificial Intelligence, Vol. 24, pp. 169-203, 1984.

Lee, W., Experimental Design and Analysis, San Francisco, CA: Freeman, 1975.

Mehle, T., "Hypothesis generation in an automotive malfunction inference task," Acta Psychologica, Vol. 52, pp. 87-106, 1982.

Miller, R.A., Pople, H.E. Jr., and Myers, J.D., "INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine," in Readings in Medical Artificial Intelligence, Clancey, W.J. and Shortliffe, E.H. (eds.), Reading, MA: Addison-Wesley, 1934.

Morris, N.M. and Rouse, W.B., "The effects of type of knowledge upon human problem solving in a process control task," IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-15, No. 6, 1985.

Morris, N.M. and Rouse, W.B., "Review and evaluation of empirical research in troubleshooting," Human Factors, Vol. 27, No. 5, October 1985.

NASA Johnson Space Center, "Orbital refueling demonstration system description," Program Development Office, October 21, 1985.

Neter, J. and Wasserman, W., Applied Linear Statistical Models, Homewood, IL: Irwin, 1974.

Newell, A. and Simon, H.A., Human Problem Solving, Englewood Cliffs, NJ: Prentice-Hall, 1972.

Wohl, J.G. "Cognitive capability versus system complexity in electronic maintenance," _IEEE Trans. Systems, Man, and Cybernetics,_ Vol. SMC-13, No. 4, 1983.

Yoon, W.C. and Hammer, J.M. "Aiding the operator during novel fault diagnosis," to appear in _IEEE Trans. Systems, Man, and Cybernetics, 1987_.

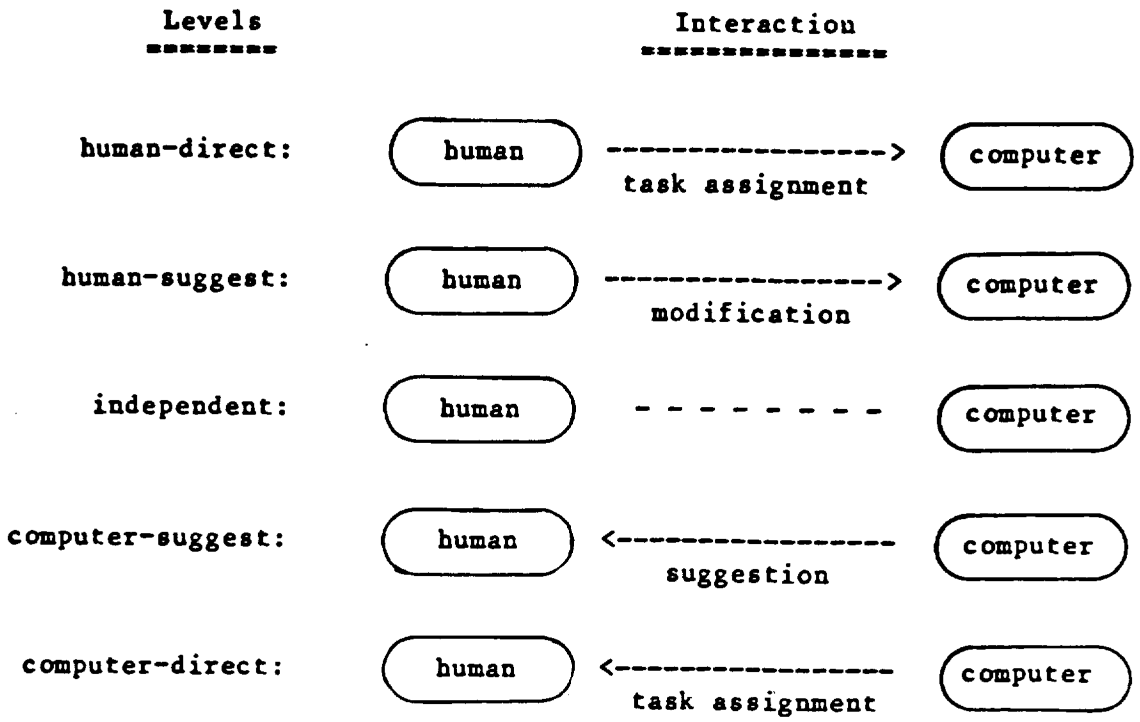


Figure 1. Levels of Interaction

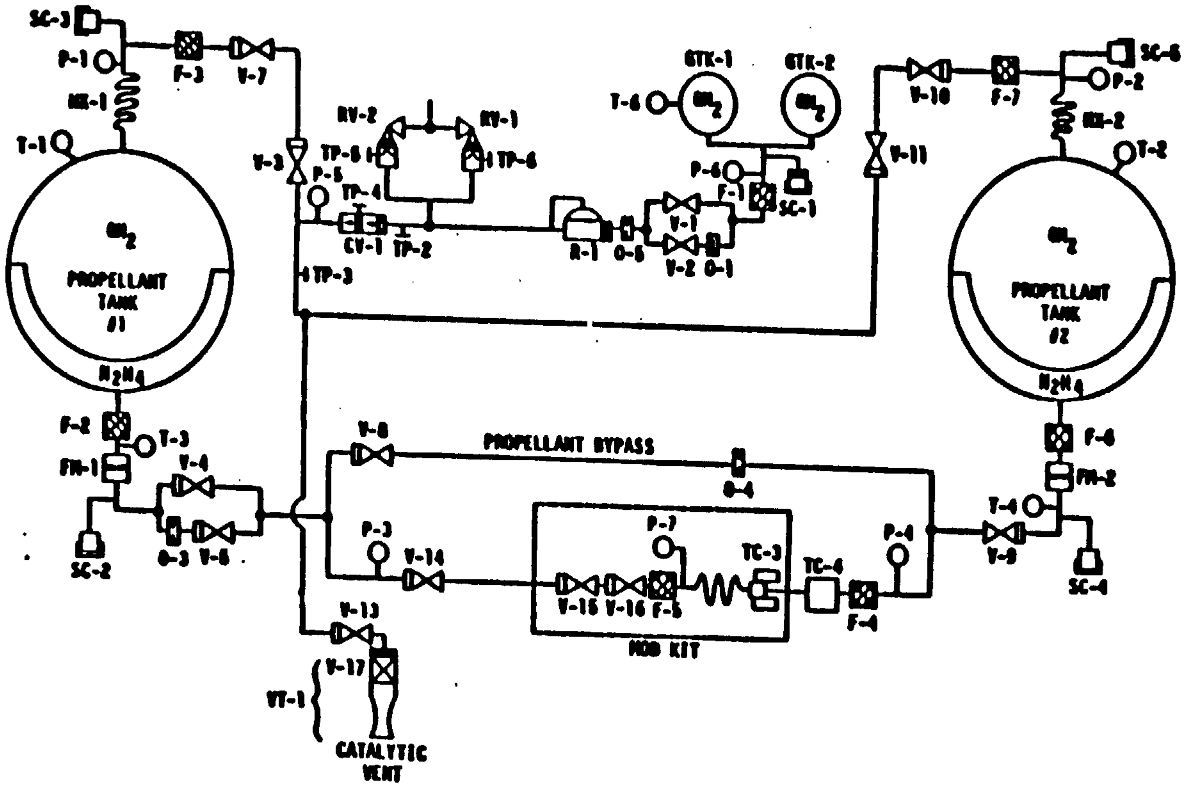


Figure 2. The Orbital Refueling System.

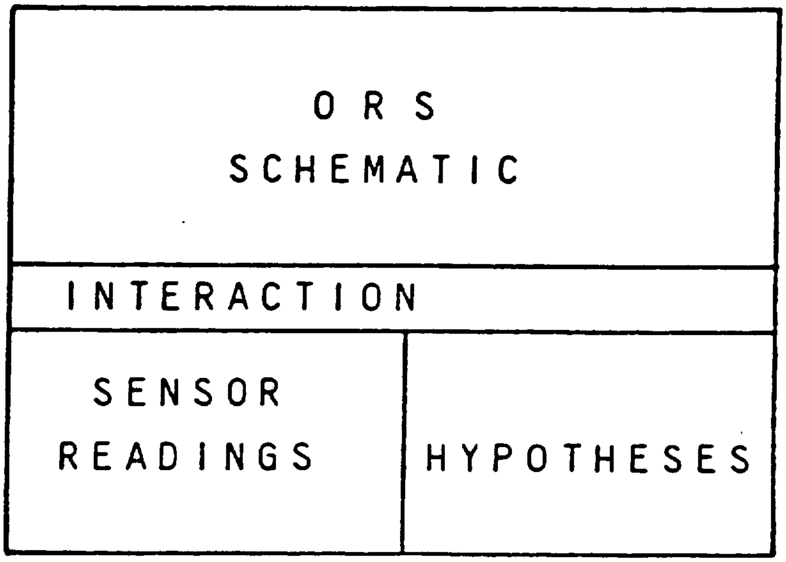


Figure 3. The operator's display.

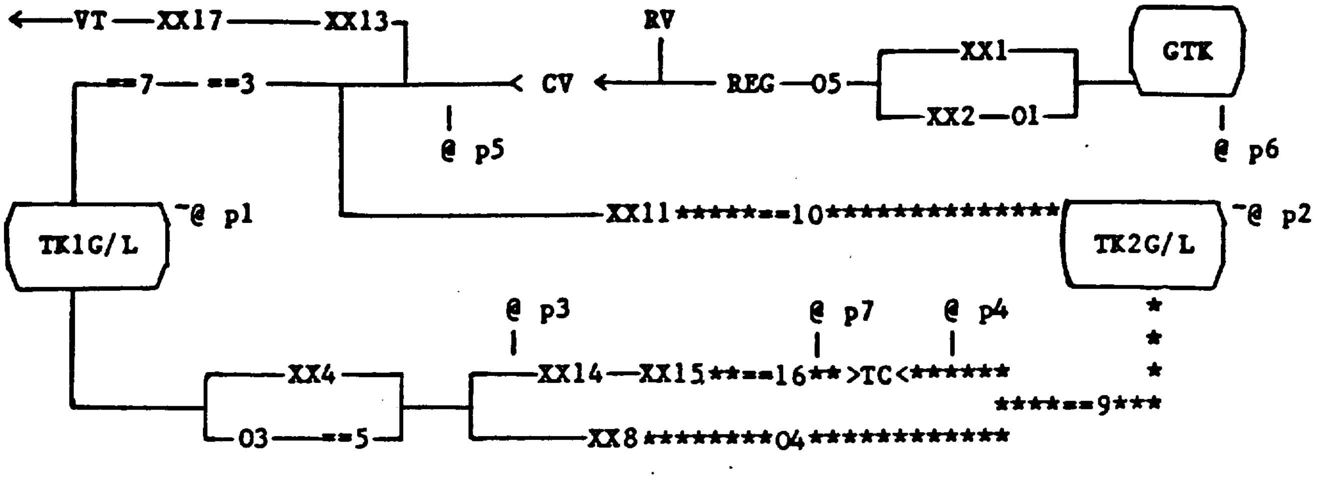


Figure 4. The normal response (N).

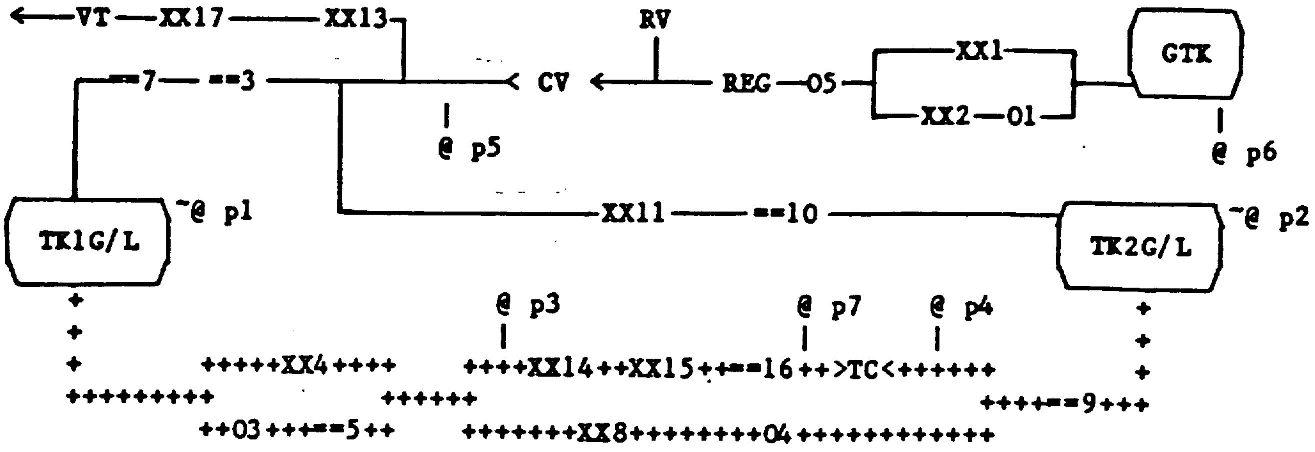
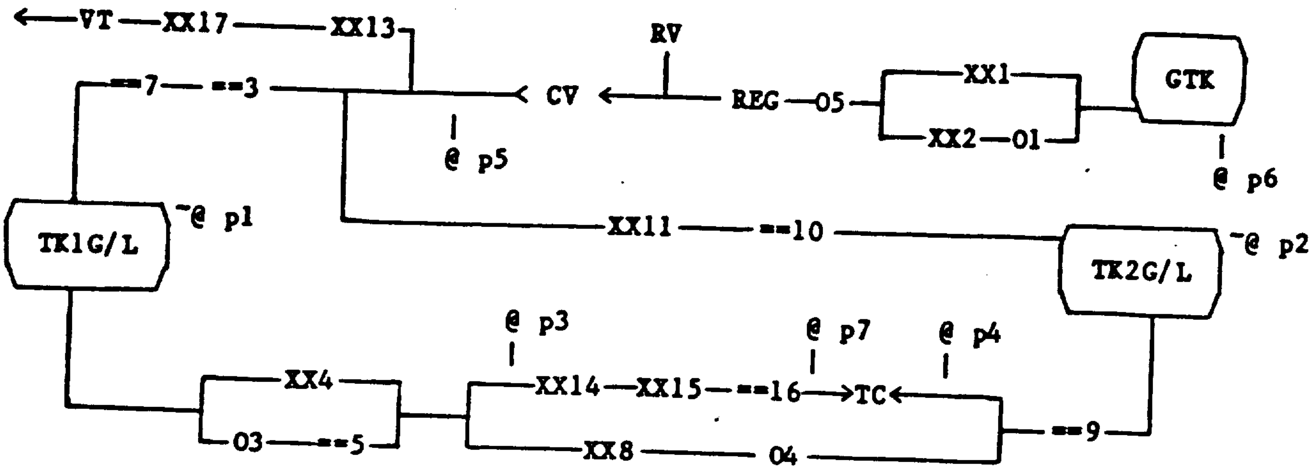
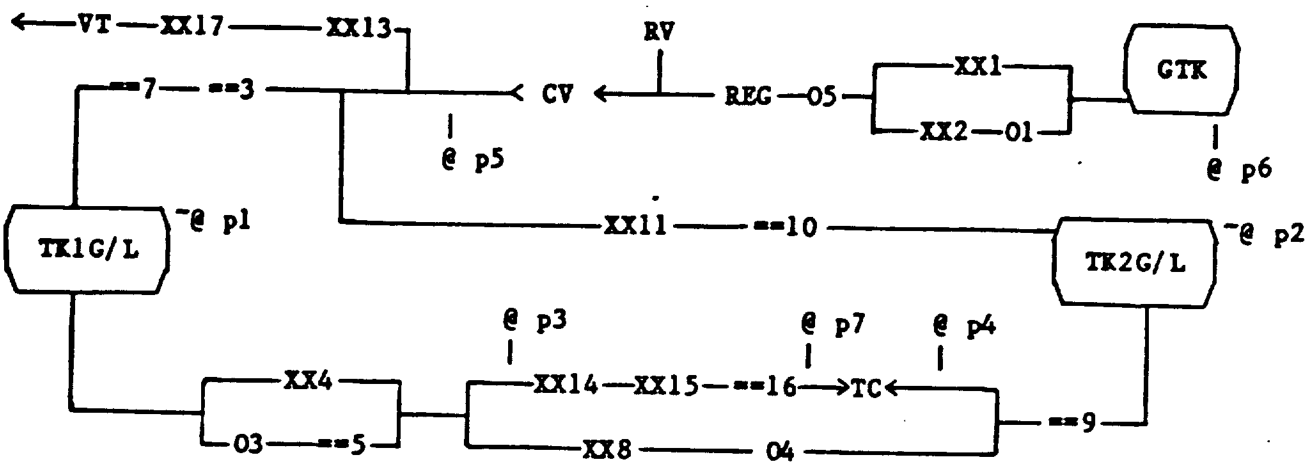


Figure 5. The observed response (0).



SEE
* D P1
* D P2
* C P2 P4

Figure 6. Deviation from normal behavior (O-N).



```
SEE
 * D P1
 * C P2 P4
```

(L V10): leak

Figure 7. Deviation from hypothesized behavior (O-H).

		PROBLEMS					
		P1	P2	P3	P4	P5	P6
GROUPS	G1 (S1-S3)	-	N	-	N	-	N
	G2 (S4-S6)	N	-	N	-	N	-

where N: N-aided situation
-: unaided situation

Figure 8. Latin Square Design for N effects in Experiment 1.

		PROBLEMS					
		P1	P2	P3	P4	P5	P6
GROUPS	G1 (S1-S3)	-	O	O-N	-	O-N	O
	G2 (S4-S6)	O	O-N	-	O	-	O-N
	G3 (S7-S9)	O-N	-	O	O-N	O	-

where O: O-aided situation
O-N: O-N aided situation
-: unaided situation

Figure 9. Latin Square Design for O and O-N in Experiment 2.

		PROBLEMS							
		P1	P2	P3	P4	P5	P6	P7	P8
GROUPS	G1(S1,S2)	-	A	-	A	-	A	-	A
	G2(S3,S4)	A	-	A	-	A	-	A	-

where A: aided situation (O-N or O-H)
-: unaided situation

Figure 10. Latin Square Designs for O-N and O-H effects in Experiment 3.

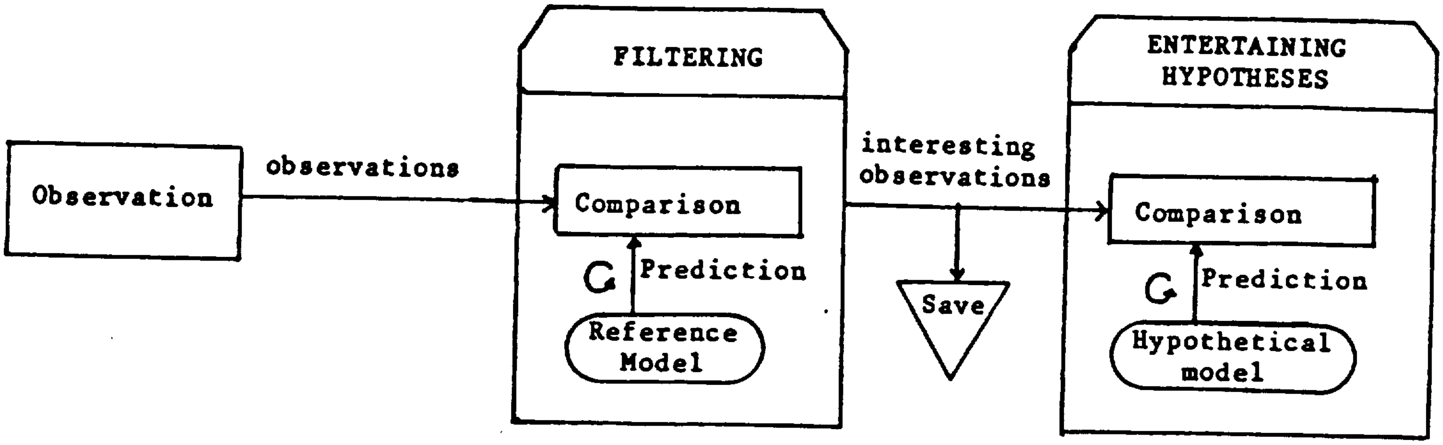


Figure 11. Data-driven Search

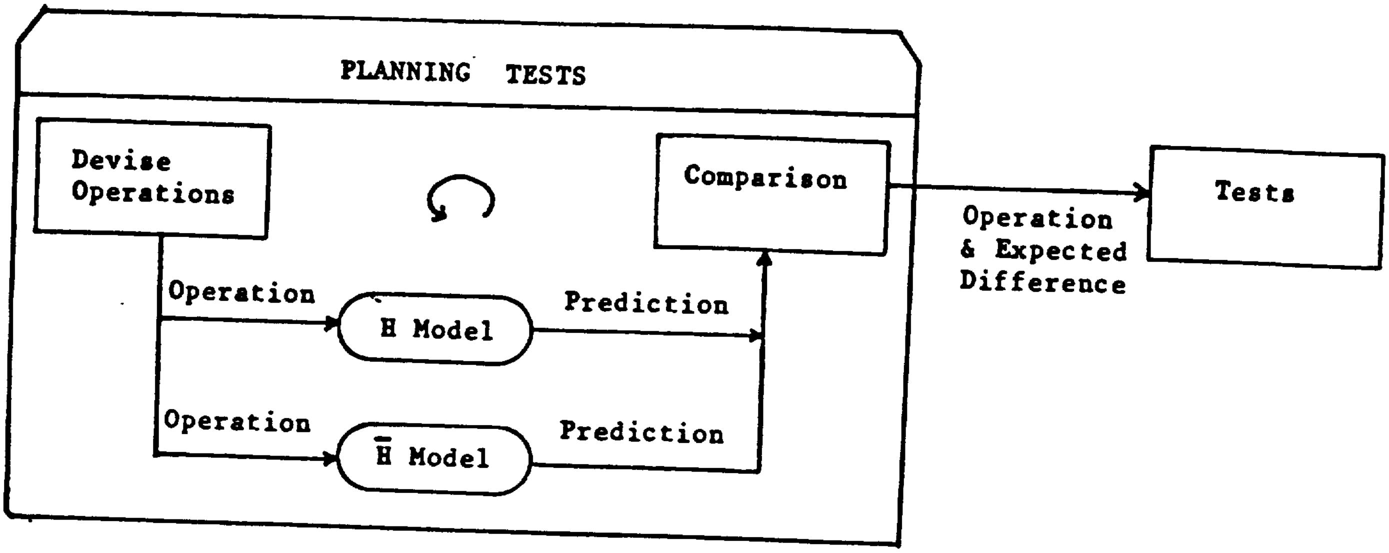


Figure 12. Hypothesis-driven Search